

# Adaption of the OMOP CDM for Rare Diseases

Michele ZOCH<sup>a,1</sup>, Christian GIERSCHNER<sup>a</sup>, Yuan PENG<sup>a</sup>, Mirko GRUHL<sup>a</sup>, Liz. A. LEUTNER<sup>a</sup>, Martin SEDLMAYR<sup>a</sup> and Franziska BATHELT<sup>a</sup>

<sup>a</sup>*Institute for Medical Informatics and Biometry at Carl Gustav Carus Faculty of Medicine at Technische Universität Dresden, Germany*

**Abstract.** The OMOP Common Data Model (OMOP CDM) is an option to store patient data and to use these in an international context. Up to now, rare diseases can only be partly described in OMOP CDM. Therefore, it is necessary to investigate which special features in the context of rare diseases (e.g. terminologies) have to be considered, how these can be included in OMOP CDM and how physicians can use the data. An interdisciplinary team developed (1) a Transition Database for Rare Diseases by mapping Orpha Code, Alpha ID, SNOMED, ICD-10-GM, ICD-10-WHO and OMOP-conform concepts; and (2) a Rare Diseases Dashboard for physicians of a German Center of Rare Diseases by using methods of user-centered design. This demonstrated how OMOP CDM can be flexibly extended for different medical issues by using independent tools for mappings and visualization. Thereby, the adaption of OMOP CDM allows for international collaboration, enables (distributed) analysis of patient data and thus it can improve the care of people with rare diseases.

**Keywords.** Rare Diseases, OMOP CDM, Interoperability, Secondary Data Use, Transition Database, Dashboard

## 1. Introduction

In Europe, a disease is defined as rare if it concerns fewer than 5 in 10,000 people. With an estimated count of 6,000 to 8,000 different rare diseases, it is assumed that around 13.5 million people in Europe suffer from a rare disease [1, 2]. The combination of the high estimated number of people affected, the variability of the diseases and the complexity of their manifestations (like chronic illnesses, physical disabilities, mental illnesses) poses a major challenge for the health care system [3]. Diagnosis and adequate therapy of rare diseases are made more difficult due to the lack of expertise on the multitude of clinical pictures as well as the geographically dispersed experts and country-specific documentation. It was determined that patients with rare diseases get diagnosed only after 5 to 30 years and approximately 69 % of patients with rare diseases are initially misdiagnosed in Germany [4].

Clinical decision support systems (CDSS) could be an option to overcome these challenges. A recent study shows that currently 19 CDSS are available to support the diagnosis and therapy of rare diseases [5], for which reliable information and data are

---

<sup>1</sup> Corresponding Author: Michele Zoch, Institute for Medical Informatics and Biometry at Carl Gustav Carus Faculty of Medicine at Technische Universität Dresden, Fetscherstraße 74, 01407 Dresden, Germany; E-mail: michele.zoch@tu-dresden.de.

needed. However, studies show the lack of a large amount of real world data in the context of rare diseases [2, 6, 7].

Another chance accelerating diagnosis is to strengthen international research and collaboration by connecting medical experts and thus also their data. This might enhance the common understanding of rare diseases, for example by using the same terminology (semantic level). Furthermore, the data could be available in a harmonized form to facilitate corporate research and distributed analysis (syntactic level).

One way to achieve the necessary standardization is to map clinical data into the *Observational Medical Outcome Partnership Common Data Model* (OMOP CDM) [8]. In the context of rare diseases, the adaption of OMOP CDM is essential due to the current lack of catalogs with the needed terminologies (e.g. Orpha Code, Alpha ID). In this paper, we demonstrate an approach, which imports the terminologies and makes use of the data to provide a dashboard for physicians in the field of rare diseases.

## 2. Methods

The proposed approach was discussed and consented in an interdisciplinary team of data scientists, computer scientists and physicians. Developing the concept, the following special characteristics of rare diseases were taken into account:

- Coding rare diseases with ICD-10 is not sufficient; in 2015, about 335 of 6.000 to 8.000 rare diseases could be documented with ICD-10 [9]. Therefore, alternative coding options, such as Orpha Codes [10] and the German Alpha ID [11], are used. Often, physicians and medical documentation officers work with multiple codes of different catalogs in order to guarantee correct accounting and the clearest possible designation.
- No restrictions are made to specific medical specialties (e.g. neurology) or domains (e.g. laboratory values, conditions), to provide an overall picture of the health status of patients.

The Vocabulary Team of OHDSI, the community that provides OMOP CDM, by including them in *ATHENA*, does the insertion of additional terminologies for the use of OMOP CDM. This tool covers the standardized, OMOP-conform vocabularies, which links the data with uniform concepts [12]. In order to use the mappings of the terminologies without dependencies of OHDSI and *ATHENA*, we developed the Transition Database for Rare Diseases. It is based on the idea of the “German Transition Database” for harmonizing German claim data and OMOP-conform concepts [13]. In the context of rare diseases, we used files from Orphadata [14] and the German Federal Institute for Drugs and Medical Devices [15] and mapped the different expression to the Orpha Number, which is unique and clearly identifies a rare disease.

To give physicians a broad view of patient data, we connected OMOP CDM to a dashboard tool. For this, we used the tool *DashBuilder* [16]. This approach is developed as an addition to *ATLAS*, which is an OHDSI tool for secondary data use through cohorting [17]. Unlike *ATLAS*, the Rare Diseases Dashboard provides an insight into the current health status of patients with defined indicators and so without defining inclusion and exclusion criteria for medical cohorts. In cooperation with the University Center for Rare Diseases in Dresden, requirements were analyzed, mockups implemented and evaluated by physicians to set up the dashboard.

### 3. Results

The Transition Database includes mappings between Orpha Code, Alpha ID, SNOMED, ICD-10-GM and ICD-10-WHO (see Figure 1). It covers about 10,500 mappings of Orpha Code, SNOMED, ICD-10-GM and ICD-10-WHO as well as about 7,200 mappings of Orpha Code and Alpha ID.

orpha_number	orpha_name	icd10who	icd10gm_primaer_1	snomed_code	orpha_number	alphaid_n
5	Langkettien-3-Hydroxyacyl-CoA-Dehydrogenase-Mangel	E71.3	E71.3	726021008		51117853
6	3-Methylcrotonyl-CoA-Carboxylase-Mangel, isolierter	E71.1	E71.1	13144005		51117842
7	3C-Syndrom	Q87.8	Q87.8	718558007		61118351
8	47,XXX-Syndrom	Q98.5	Q98.5	50749006		61118175
9	Tetrasomie X	Q97.1	Q97.1	10567003		61118027
10	48,XXX-Syndrom	Q98.8	Q98.8	403760006		71115654
11	Pentasomie X	Q97.1	Q97.1	432480007		71115653
13	6-Pyruvoyl-Tetrahydropterin-Synthase-Mangel	E70.1	E70.1	(null)		81124884
14	Abetalipoproteinämie	E78.6	E78.6	150787008		819268
15	Achondroplase	Q77.4	Q77.4	86268005		9179459
16	Blauzapfenmonochromasie	H53.5	H53.5	789676005		91120048
17	Fatale infantile Laktatazidose mit Methylmalonazidurie	E71.1	E71.1	715338007		9179425

Figure 1. Screenshot of the Transition Database; left side: mapping of Orpha Number, ICD-10-WHO, ICD-10-GM and SNOMED, right side: mapping of Orpha Number and Alpha-ID

Figure 2 shows a screenshot of the Rare Diseases Dashboard. It present indicators like the number of medical cases with rare diseases by diagnosis status (confirmed, tentative, excluded) or number of medical cases per number of rare diseases.

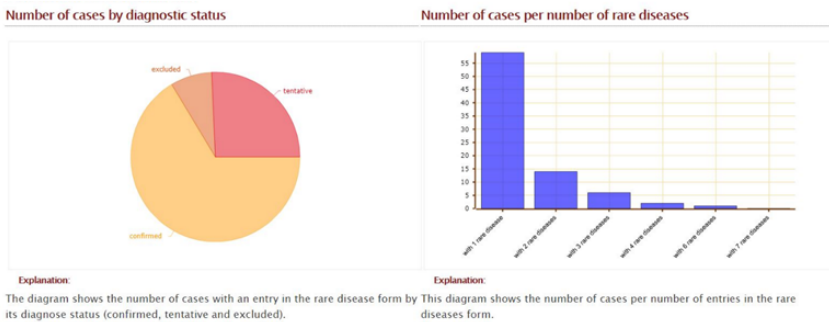


Figure 2. Screenshot of the Rare Diseases Dashboard (based on synthetic data)

### 4. Discussion

The concept described is limited by German specifics, such as the documentation of rare diseases with the German Alpha-ID. Nevertheless, international terminologies such as SNOMED and Orpha Code are also included. Furthermore, the quality of the mappings depends on the quality of source files developed, quality assured and provided by different institutes for medical research (e.g. the German Federal Institute for Drugs and Medical Devices). A further quality assurance by physicians or by means of medical ontologies or metathesaurus (e.g. Unified Medical Language System) are still pending.

The Transition Database is productively used to specify mappings of German terminologies. We are using it to ensure semantic and syntactic interoperability between the clinical data warehouse and different research data repositories like OMOP. By applying mappings to SNOMED, the terminologies could be mapped to a standardized vocabulary and thus utilized in OMOP CDM. This demonstrates the extensibility of the CDM to further areas of medical research. Extensions of the standard OMOP vocabulary

with national terminologies or terminologies for specific medical domains are common in the OHDSI environment, as recent studies show [18, 19]. The described approach enables the storage and maintainability of mappings independent of *ATHENA*. The combination of a custom dashboard and OMOP CDM shows that the CDM can be used as a basis of other visualization approaches independent of *ATLAS*. It is also conceivable that other dashboard tools can be connected, like e.g. *Superset* [20].

The aforementioned characteristics of rare diseases are taken into account by mapping the terminologies and transferring them to OMOP CDM and by presenting the relevant data within a dashboard. An evaluation of the entire concept as well as a proof-of-concept will be done in the future.

## 5. Conclusion

The presented approach shows possibilities for adapting OMOP CDM in the context of rare diseases: Both an extension with specific terminologies and a connection to a dashboard for the visualization of indicators for a German Center for Rare Diseases were demonstrated. It is shown that an external database for storing transitions is also usable outside of the OHDSI community (e.g. see [13]). Thus, while the mappings can be used for OMOP CDM, they can also be used independently for other use cases, such as in the clinical data warehouse. The interaction of OMOP CDM and the Transition Database forms the basis for interoperable data on a semantic and syntactic level. This can empower secondary data use, e.g., for CDSS as well as for international research and distributed analysis. The connection to a dashboard shows the applicability of the CDM to individual medical questions.

Future work will investigate how the Transition Database can be expanded: (1) when a rare disease is suspected but cannot yet be designated, the Human Phenotype Ontology (HPO) can be included to code phenotypic abnormalities [21]. (2) In combination to the documentation of phenotypes, the storage of genetic data is important as well, because about 80 % of rare diseases are genetic [2]. For this purpose, it is possible to add four additional tables under the name “OMOP G-CDM” [22].

Furthermore, it is necessary to check options to include additional data like imaging data; possibly through the connection of *XNAT* [23], a platform to manage and explore medical imaging. Also it will be scrutinized how OMOP CDM could be used to give a more detailed insight in the current health status of patients with rare diseases by visualizing the treatment pathway. In addition to empower collaborative research, it should be reviewed how distributed analyses can be run based on the proposed concept; e.g. with the help of *DataShield* [24]. The connection between the mentioned technologies and OMOP CDM forms the prerequisite for the better insight into rare diseases in general and the status of patients with such illnesses in particular.

## 6. Acknowledgement

This work is part of the projects MIRACUM (FKZ 01ZZ1801A/L) and CORD (FKZ 01ZZ1911I), funded by the German Ministry of Education and Research.

## References

- [1] Aymé S, Schmidtké J. Networking for rare diseases: a necessity for Europe. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*. 2007;50(12):1477–1483.
- [2] Schieppati A, Henter JL, Daina E, Aperia A. Why rare diseases are an important medical and social issue. *The Lancet*. 2008;371(9629):2039–2041.
- [3] Dharsri S, Wong-Rieger D, Harold M, Terry S. Review of 11 national policies for rare diseases in the context of key patient needs. *Orphanet Journal of Rare Diseases*. 2017;12(1):63.
- [4] Eurordis. The voice of 12,000 patients: experiences and expectations of rare disease patients on diagnosis and care in Europe : a report based on the EurordisCare3 surveys. Eurordis; 2009, Available at: [https://www.eurordis.org/IMG/pdf/voice\\_12000\\_patients/EURORDISCARE\\_FULLBOOKr.pdf](https://www.eurordis.org/IMG/pdf/voice_12000_patients/EURORDISCARE_FULLBOOKr.pdf).
- [5] Schaaf J, Sedlmayr M, Schaefer J, Storf H. Diagnosis of Rare Diseases: a scoping review of clinical decision support systems. *Orphanet Journal of Rare Diseases*. 2020;15(1).
- [6] Schaefer J, Lehne M, Schepers J, Prasser F, Thun S. The use of machine learning in rare diseases: a scoping review. *Orphanet Journal of Rare Diseases*. 2020;15(1).
- [7] Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. *Journal of the American Medical Informatics Association*. 2015;22(1):76–85.
- [8] OHDSI. OMOP Common Data Model; 2021, Available at <https://www.ohdsi.org/data-standardization/the-common-data-model/> Accessed January 9, 2021.
- [9] Ayme S, Bellet B, Rath A. Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. *Orphanet Journal of Rare Diseases*. 2015;10(1).
- [10] Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. *Human Mutation*. 2012;33(5):803–808.
- [11] Weber S, Dávila M. German approach of coding rare diseases with ICD-10-GM and Orpha numbers in routine settings. *Orphanet Journal of Rare Diseases*. 2014;9(1).
- [12] OHDSI. ATHENA standardized vocabularies; 2021, Available at <https://www.ohdsi.org/analytic-tools/athena-standardized-vocabularies/>, Accessed January 9, 2021.
- [13] Kümmel M, Reinecke I, Gruhl M, Bathelt F, Sedlmayr M. Transition Database for a harmonized mapping of German patient data to the OMOP CDM; 2020, Available at <https://www.ohdsi.org/2020-eu-symposium-showcase-13/>, Accessed January 9, 2021.
- [14] Orphadata, Available at <http://www.orphadata.org>.
- [15] Alpha-ID, Available at <https://www.dimdi.de/dynamic/en/classifications/icd/alpha-id/index.html>.
- [16] Dashbuilder, <http://dashbuilder.org/>.
- [17] OHDSI. ATLAS – A unified interface for the OHDSI tools; 2021, Available at <https://www.ohdsi.org/atlas-a-unified-interface-for-the-ohdsi-tools/>, Accessed January 9, 2021.
- [18] Warner JL, Dymshyts D, Reich CG, Gurley MJ, Hochheiser H, Moldwin ZH, Belenkaya R, Williams AE, Yang PC. HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. *J Biomed Inform*. 2019 Aug;96:103239.
- [19] Zhang Y, Li J, Zandt MV. NCCD-RxNorm: Linking Chinese Clinical Drugs to International Drug Vocabulary. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2020. p. 1752–1756.
- [20] Apache Superset, Available at <https://superset.apache.org/>.
- [21] Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. *The American Journal of Human Genetics*. 2008;83(5):610–615.
- [22] Shin SJ, You SC, Park YR, Roh J, Kim JH, Haam S, et al. Genomic Common Data Model for Seamless Interoperation of Biomedical Data in Clinical Practice: Retrospective Study. *Journal of Medical Internet Research*. 2019;21(3).
- [23] XNAT, Available at <https://www.xnat.org/>.
- [24] Horki P, Lenz S, Gruendner J, Maier C, Liebler A, Boeker M. omopRds: transfer of data models from OMOP to DataSHIELD/Opal. German Medical Science GMS Publishing House; 2019.